

PoNTE: apontando para corpos de aprendizes de tradução avançados

PoNTE: a bridge for advanced translation learners

Diana Santos

Linguatca, Universidade de Oslo

d.s.m.santos@ilos.uio.no

Resumo

É possível ensinar usando os materiais criados pelos próprios alunos, e ao mesmo tempo anotá-los para obter mais material que fique público para mais professores e estudantes? É possível desenvolver o DISPARA, inicialmente concebido para disponibilizar corpos de tradução “tradicionais”, de forma a conter mais um nível de anotação de “crítica de tradução”? O projeto PoNTE pretende ser uma abordagem inicial a estas duas questões.

Neste artigo, descrevo o tipo de comentários e anotação crítica que seria desejável ter codificado num corpo deste género, a sua primeira implementação e estudos realizados, e os problemas técnicos que ainda se põem na gestão de um corpo sempre crescente.

Palavras chave

Ensino de português como língua estrangeira, corpos paralelos, ferramentas de apoio ao ensino, ferramentas de corpos, tradução, linguística contrastiva

Abstract

Is it possible to teach using materials created by the students themselves, while at the same time annotate them to raise more material available to other teachers and students? Is it possible to further develop DISPARA, initially deployed for “traditional” translation corpora, so that it contains an extra annotation level with “translation critique”? The PoNTE project (“*ponte*” means bridge in Portuguese, and PoNTE stands for “Portuguese-Norwegian Translation Examples”) offers an initial approach to these two questions.

In this paper, I describe the type of comments and critical annotations aimed for, together with a first implementation, a short reference to some studies performed, and discuss the technical and philosophical problems involved in the management of a dynamic and always growing corpus.

Keywords

Portuguese for foreigners, parallel corpora, computer aided language learning, teaching tools, corpora tools, translation, contrastive linguistics

1 Apresentação

A esmagadora maioria dos corpos paralelos envolvendo o português tem o inglês como A outra língua, ou como a língua pivô, o que se explica naturalmente pelo peso que o inglês tem a nível internacional e o facto de corresponder à língua mais traduzida para português e vice-versa (a língua na qual se encontram mais textos traduzidos de um original em português, cf. Rosa (2006)).

Contudo, e sobretudo numa situação de ensino da língua, é conhecida a importância da língua materna assim como a importância do treino na tradução e na retroversão associada à língua materna. (Estou plenamente consciente de que existem várias opiniões divergentes sobre o assunto, mas eu alinho com a fação, possivelmente minoritária presentemente, que afirma que esse é um assunto que apaixona os alunos como poucos, em completo acordo com o que escreve Kåre Nilsson (1997), o meu predecessor na Universidade de Oslo.)

Na Universidade de Oslo, a esmagadora maioria dos alunos tem o norueguês como língua materna, e os poucos que têm o português ou outra têm excelentes conhecimentos do norueguês, o que tornou óbvio um ensino da gramática (e cultura) relacionado – também – com a tradução. Assim surgiu o PoNTE (Portuguese-Norwegian Translation Examples), um corpo de múltiplas traduções entre as línguas portuguesa e norueguesa, que usa as traduções dos alunos como textos de chegada.

Ao contrário do que é geralmente assumido pelos pessoas de fora do ambiente do ensino das línguas, os professores de línguas na sua prática pedagógica não têm geralmente tempo para criar ou recorrer a um corpo criado a partir dos materiais dos seus próprios alunos. Quando criam esses corpos, é com a ideia de que virão um dia a ser úteis na sua prática diária, ou para os tempos vindouros.

Um bom indicador do irrealismo desta esperança foi o recente inquérito de Sylviane Granger na lista corpora¹, a 22 de outubro de 2013:

I'm looking for as much information as possible on concrete uses of learner corpus data for: (...) Many publications describe the potential use of learner corpus data, but I would like to collect information on actual use.²

Sylviane Granger é talvez a mais famosa especialista de corpos de aprendentes, como se pode ver por Granger, Gilquin e Meunier (2013). O seu pedido na lista teve resposta de nove investigadores diferentes (sumarizado a 31 de dezembro de 2013), e apenas um semelhante ao PoNTE (Kutuzov et al., 2012), que discutiremos mais adiante.

Por essa razão, pode ser interessante realçar que praticamente desde o primeiro instante eu usei o PoNTE no meu ensino, ou melhor, os resultados da invocação da sua primeira versão, como auxiliar pedagógico na sala de aula.

De qualquer forma, não me distingo dos outros investigadores quando reconheço que a criação última deste corpo pretende vir a permitir – mais tarde, quando tiver chegado a um maior tamanho, ou no caminho enquanto é incrementado – identificar, e permitir estudar:

- o que é difícil de compreender em português para alunos noruegueses,
- o que é difícil de traduzir para norueguês, e
- o que é difícil de exprimir em português (vindo do norueguês).

É importante realçar, neste contexto, que a própria construção do corpo e a simples comparação das traduções por si só não é necessariamente útil para os “tradutores” (alunos de bacharelato), que podem até ser inundados de piores traduções do que as suas próprias. Como sempre insisti ao apresentar o material, e eles também pedem, o mais importante para a sua aprendizagem é o comentário por parte do professor.

Assim, em 2012 passei a disponibilizar, aos tradutores de cada texto, o conjunto das traduções, marcando a negrito os casos em que

tinha havido clara diferença entre as traduções, e nesse caso escolhendo o trecho (que pode ir de uma palavra à unidade de alinhamento total) que me parecia melhor. Nos casos de nenhuma tradução ser satisfatória, marquei a negrito o (trecho do) original.

Este material serve assim como lembrança da aula em que estiveram e ouviram os meus comentários, mas naturalmente não permite a subsequente recuperação, por parte de outro utilizador do corpo, dos problemas e soluções anotados, até porque apenas sublinha partes do texto.

Por isso a nossa intenção de anotar o PoNTE com informação mais explícita e estruturada para subsequente procura, informação que chamo, a partir daqui, “informação crítica”.

Embora o PoNTE seja um trabalho em progresso, porque a anotação crítica ainda não foi incorporada nem fixado o sistema subjacente para essa mesma anotação, a mera existência de um corpo paralelo nas duas línguas, com abundância de traduções de um mesmo texto, já permitiu e permite alguma investigação interessante, como tentaremos demonstrar no presente artigo.

Assim, após descrever brevemente o conteúdo presente na secção 2 e a implementação atual na secção 3, faço uma breve menção a alguns estudos já efetuados sobre (versões anteriores de) o PoNTE na secção 4 e discuto finalmente a questão da anotação crítica na secção 5, documentando brevemente também duas anotações piloto.

A publicação deste texto tem, assim, dois objetivos: anunciar e documentar um novo recurso, e explicar as opções tomadas e as que ainda estão por tomar.

2 Breve descrição do conteúdo do corpo

O PoNTE é um corpo em permanente evolução visto que deriva de uma atividade didática continuada.

Os originais são textos curtos, selecionados para serem traduzidos, por vezes correspondendo a excertos de textos. Dado que as aulas têm lugar todos os anos, prevê-se que o número de textos diferentes, assim como o número das suas traduções, vá aumentando a um ritmo constante.

Em janeiro de 2014, os textos presentes no corpo, assim como o número de traduções distintas, estão indicados na Tabela 1 para originais em português e na Tabela 2 para originais em norueguês.

¹Lista de discussão corpora@uib.no

²A minha tradução: Estou à procura de exemplos concretos de corpos de aprendentes para : (...) Muitas publicações descrevem o uso potencial de corpos de aprendizes, mas eu gostava de recolher informação sobre o uso real.

A lista das fontes encontra-se acessível do sítio do projeto³.

Os textos são de variantes diferentes, de géneros diferentes, e publicados em suportes diferentes, só tendo em comum possuírem um tamanho razoavelmente pequeno (ver Figura 1 para o número de palavras).

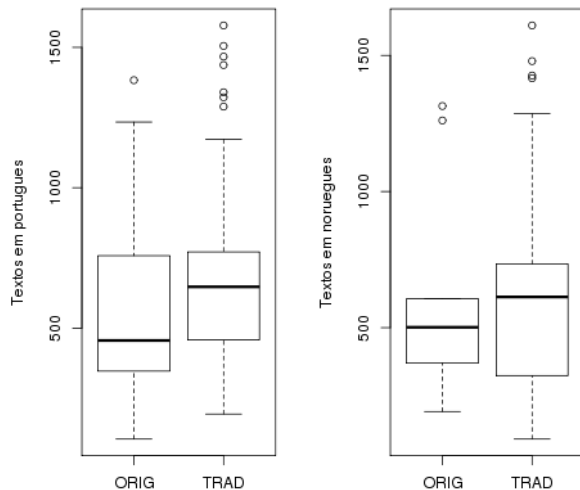


Figura 1: Tamanho dos textos do PoNTE em número de palavras. Repare-se que TRAD se refere às traduções dos ORIG da outra língua.

Assim, e sem entrar em pormenores, temos textos brasileiros, moçambicanos, portugueses e angolanos, e contos, reportagens, blogs, humor, burocracia, promessas eleitorais e texto técnico no que toca aos originais em português, e textos jornalísticos, literários e cartas no que se refere ao norueguês, que também é representado em três variantes (bokmaal, nynorsk e riksmaal⁴).

De uma forma mais condensada, podemos dizer que existem 266 pares de tradução, 195 traduções para norueguês e 71 para português. A razão desse desequilíbrio é clara: há muito mais alunos nas cadeiras iniciais (que traduzem do português para o norueguês) do que nas mais avançadas, em que fazem retroversão para português.

Mais importante do que a constituição presente, contudo, é a análise das potencialidades deste material para o ensino da tradução e para a linguística contrastiva, que serão tornadas possíveis pela sua anotação crítica. Depois de

³<http://www.linguateca.pt/PoNTE/>

⁴Para quem não conheça a situação linguística da Noruega, menciono simplesmente que existem duas grafias modernas, e uma antiga, próxima do dinamarquês atual.

Texto	Trads	Formas	Tipos
AMAZ	10	161	113
BP	5	276	162
BRI	16	234	142
CAMP	9	456	233
CAR	18	646	299
CIE	16	1234	454
DDS	3	687	272
DIL	11	958	436
DSC	17	795	400
EDS	6	305	156
ELEI	7	734	322
EPA	4	1383	525
EXA	9	105	70
JP	4	346	196
LOG	2	417	202
MEC	1	857	306
MIA	5	834	403
MRC	14	758	326
MUL	3	448	264
OC	18	746	346
PIB	2	379	190
SEM	6	371	212
SPG	1	557	270
TED	6	368	209
VAN	2	347	156

Tabela 1: Composição atual do PoNTE, em termos dos originais em português.

Texto	Trads	Formas	Tipos
BEB	2	1314	501
BRS	9	523	262
CLI	11	192	130
DN	3	362	222
KB	2	370	222
MOB	13	480	258
MUS	3	389	224
QUEI	8	1261	450
SAU	14	606	324
VES	6	570	282

Tabela 2: Composição atual do PoNTE, em termos dos originais em norueguês.

descrever o estado atual do projeto, e a sua implementação, debruçar-nos-emos sobre esta.

3 Implementação

Como seria natural, a implementação do PoNTE é mais uma adaptação do DISPARA (Santos, 2002) originalmente criada para o COMPARA (Frankenberg-Garcia e Santos, 2002), mas que tem sido expandida e modificada para outros corpos paralelos como o Squirrel (Borin, Carlson e Santos, 2001), o CorTrad (Tagnin, Teixeira e

Santos, 2009; Teixeira, Santos e Tagnin, 2011), e agora o PoNTE e o PANTERA⁵.

3.1 Interface

De facto, e do ponto de vista do utilizador, o PoNTE tem duas interfaces diferentes: uma que, a cada texto original, alinha todas as traduções existentes (e que é constituída por casos vazios quando o número de traduções é inferior ao máximo existente), e outra que alinha tantas vezes quantas um texto original foi traduzido, contendo assim o número total de pares original-tradução.

A razão da existência destas duas interfaces é que um utilizador pode estar interessado na comparação de traduções de um mesmo texto (caso em que usará a primeira interface), ou na contabilização de fenómenos de tradução efetuados por tradutores independentes (e, para isso, a segunda é mais natural).

Noto que a ordem das traduções não é relevante no PoNTE. Por exemplo, a segunda tradução corresponde a tradutores diferentes em textos diferentes (a numeração foi feita pela ordem de chegada das traduções, e nem todos os alunos as fizeram, nem as enviaram pela mesma ordem). Por razões de anonimização não é possível saber quem efetuou qual tradução, embora eu tenha alguma meta-informação sobre os tradutores, por exemplo a sua língua materna e a variante de português que privilegiam. A cada tradutor/aluno é atribuído um identificador, o que pode ser importante para a compreensão de alguns problemas, e esse identificador (que permitirá recuperar a língua materna, etc.) poderá vir a ser adicionado numa futura versão para refinar as procuras.

Nas figuras 2 e 3 apresenta-se a forma das duas interfaces, enquanto que as figuras 4 e 5 mostram os respetivos tipos de resultado.

3.2 Conceção teórica

De uma forma mais teórica, podemos reapresentar a implementação assim: Embora se conceba o PoNTE, abstratamente, como um único corpo, na prática ele é codificado através de uma série de diferentes corpos CWB, como aliás é ou foi o caso na implementação do COMPARA e do CorTrad, cujas interfaces invocam de facto vários corpos diferentes – conforme a direção da procura, no COMPARA, e também conforme o género, no CorTrad.

⁵Veja-se <http://www.linguateca.pt/PANTERA/> para este último, em fase de arranque.

No entanto, a conceção do PoNTE, além de também distinguir a direção, foi mais radical porque, além de conter um corpo “simples” por cada texto original, engloba também um corpo com as múltiplas traduções de cada texto, o chamado “PoNTE condensado”, e um corpo com todos os pares de traduções concebidos como um novo caso, o chamado “PoNTE distribuído”. (Este caso é semelhante aos dois únicos exemplos de traduções múltiplas no COMPARA).

Nesse aspeto, assemelha-se ao Águia (Santos, 2003), sistema de procura na Floresta Sintática, que, numa mesma página de interface na rede, junta dois corpos organizados de forma completamente diferente: um por texto, e outro por classificação gramatical (tipo de sintagma).

Ao adicionar um novo par de tradução ao PoNTE, diversos programas em Perl são ativados e (re)criam vários corpos.

3.3 Integração com o resto dos corpos da Linguateca

Outra questão extremamente importante, em que nunca é demais insistir, é o facto de que os textos em português recebem o mesmo tratamento, em termos de anotação e de revisão, que o já extenso material corpóreo da Linguateca contém: assim, tanto a anotação pelo PALAVRAS (Bick, 2000), como a anotação com a cor e outros campos semânticos, assim como a sua revisão – para um exemplo desta, cf. Maia e Santos (2012).

A possibilidade de comparar com, evocando, muito maiores quantidades de texto em português é assim algo que é oferecido automaticamente pela integração do DISPARA com o AC/DC (Costa, Santos e Rocha, 2009).

4 Alguns estudos usando o PoNTE

Aqui apresentamos brevemente dois trabalhos que já fizeram uso do PoNTE – usando diferentes versões, correspondendo a diferentes estágios do seu desenvolvimento, para motivar o seu uso e para mostrar que é possível já tirar algum proveito da sua existência, mesmo em fases preliminares de desenvolvimento.

4.1 Dativos possessivos

Um fenómeno relativamente frequente em português, e que tem merecido pouco interesse a nível monolingue, é a questão do dativo possessivo, que se torna contudo imediatamente relevante se considerarmos o seu ensino a falantes de línguas germânicas (como o norueguês ou

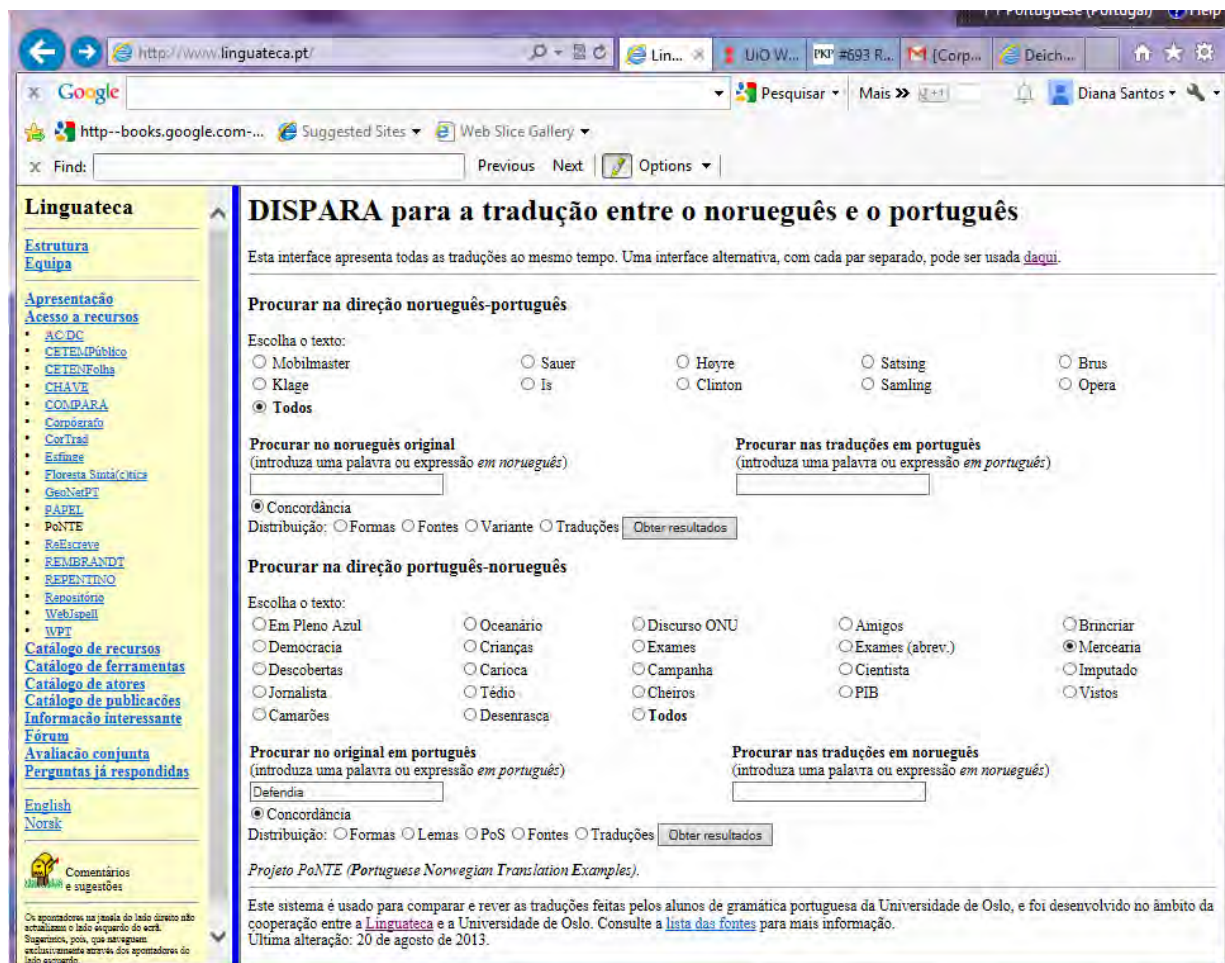


Figura 2: Exemplo de invocação do PoNTE compacto.

o inglês), devido à quase sistemática diferença entre as línguas em relação à descrição dessas situações.

Veja-se Santos (2012; Santos (2014d) para a descrição do fenómeno e da sua extensão contrastiva, usando corpos paralelos com o inglês.

Dado que o PoNTE é muito mais pequeno, que o fenómeno pressupõe um comando nativo ou quase nativo da língua, e que os aprendizes de tradutores tendem a manter a estrutura do original e não procurar traduções mais idiomáticas, a previsão que fazemos é a de que os tradutores aprendizes noruegueses muito raramente usarão dativos possessivos em português, e que, se aparecerem no original, dado ainda não terem sido alertados para esta diferença, talvez os traduzam literalmente (ou seja, como um pronome pessoal dativo).

Contudo, conseguimos identificar um caso destes no PoNTE (na tradução de português para norueguês) em que o aluno usou um pronome possessivo na tradução.

- (1) Exceto que você não estará «estudando»; você estará trabalhando, gerando conhecimento, e contribuindo para as universidades publicarem os artigos científicos que **lhes** servem como base de avaliação no cenário mundial.

Bortsett fra at du ikke skal «studere»; du skal arbeide, skaffe deg kunnskap og bidra til universitetenes publikasjoner av vitenskapelige artikler som legger grunnlaget for **deres** verdsettelse i verdenssamfunnet. ... a avaliação delas no mundo.

Este é, além disso, um caso interessante precisamente porque foi identificado em português do Brasil, onde tal fenómeno é mais raro, devido entre outras coisas à tendência para eliminação dos clíticos, bem patente nesta variante do português (Bakkejord, 2008).

O importante a reter deste caso, contudo, é que a existência de dados contrastivos para este par de línguas permite reforçar a convicção de que este assunto deve ser ensinado a alunos de português como língua estrangeira.

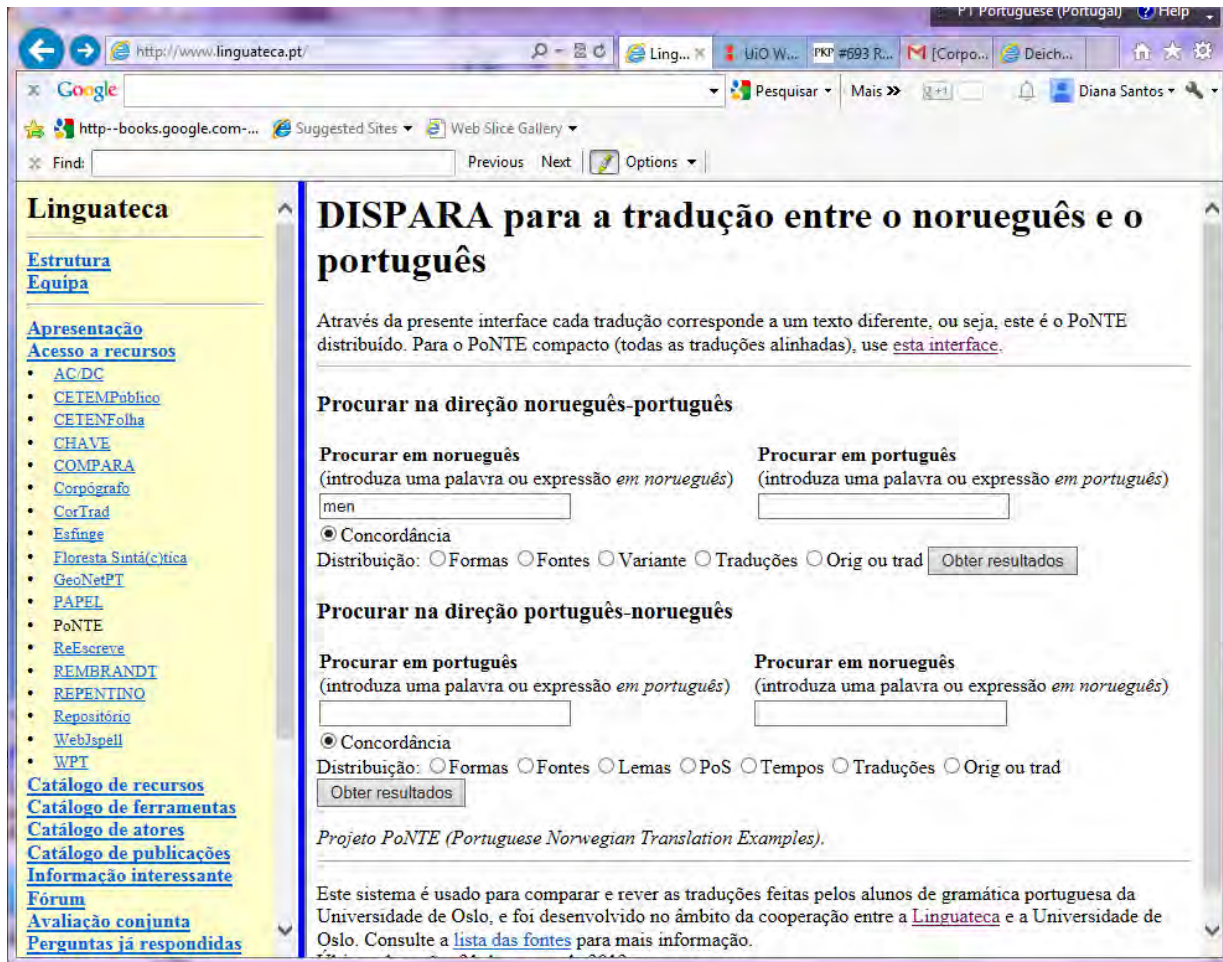


Figura 3: Exemplo de invocação do PoNTE distribuído.

4.2 Comparação de verbos estativos

Outra área em que o PoNTE já foi utilizado, agora como a principal fonte de dados, é o contraste dos verbos leves e frequentes como *ser*, *estar*, *ficar*, *ter* e *haver* (Santos, 2013; Santos, 2014e) com os seus correspondentes em norueguês.

Visto que estes verbos são dos mais frequentes nas duas línguas, além da identificação de exemplos interessantes também foi possível fazer uso da comparação de diversas traduções, e distinguir diferentes perfis entre texto traduzido e texto original, como a Figura 6 de (Santos, 2014e) ilustra.

Podemos assim afirmar, com base no PoNTE, que os textos em português têm mais menções que são traduzidas por *bli* e *være* do que os originais em norueguês.

5 Anotação crítica

Seja como for, considero que o mais importante do PoNTE ainda não foi implementado, e

que é preciso alguma reflexão em torno das funcionalidades desejáveis. Este artigo é, pois, uma forma de partilhar essas dúvidas e questões com a comunidade científica, e talvez com a comunidade de futuros utilizadores do recurso ou de recursos semelhantes.

Em primeiro lugar, **o que** pretendemos anotar?

Há dois tipos de anotação que se referem a níveis distintos da crítica de tradução, correspondentes às bem conhecidas dimensões de preservação do sentido e de fluência do texto final:

- marcar aquilo que não foi compreendido pelos tradutores (e portanto erro de conteúdo);
- apontar aquilo que foi mal expresso para o público alvo (e portanto erros de adequação ou de formulação).

A pergunta seguinte é **onde** anotar essas questões. Do lado da língua fonte, ou do lado da língua alvo?

Linguateca

- Estrutura
- Equipa
- Apresentação
- Acesso a recursos
 - AG/DC
 - CETEMPúblico
 - CETENFolha
 - CHAVE
 - COMPARA
 - Corpógrafo
 - CorTrad
 - Esfinge
 - Floresta Sintá/crítica
 - GeoNetPT
 - PAPEL
 - PoNTE
 - ReEscriere
 - REMBRANDT
 - REPENTINO
 - Repositório
 - WebSpell
 - WPT
- Catálogo de recursos
- Catálogo de ferramentas
- Catálogo de atores
- Catálogo de publicações
- Informação interessante
- Fórum
- Avaliação conjunta
- Perguntas já respondidas

Resultados da pesquisa

Eis os resultados da busca efetuada no PoNTE.

Procura: "kom.*".
distribuição de , concordância em contexto.
Corpus: PONTE-NP-PONTENC

21 ocorrências.

Concordância

Procura: "kom.*".

Num	Original	Traduções	Comentários	Gerais
1	Stein: Ehh - (ser ned) ehh (rynker pannen) ehh... der kommer bussen min (Stein skynder seg til bussen).	Stein (sic): Ehh -- (olha para baixo) ehh (enrug a testa) ehh... ali vem o meu autocarro (Stein precipita-se para o autocarro). Stein: -- Hum... (olha para baixo) hum...		
2	En kultur jeg er komfortabel med, og glad for å ta del i.	Uma cultura em que me sinto confortável, e feliz por lhe pertencer. Eu vejo isto como um resultado da cultura, uma cultura na qual me sinto confortável e fico feliz em fazer parte.		
3	Bestiller du appelsinjuice kommer du med en klar melding.	Encomendando sumo de laranja, mandamos uma mensagem distinta. Se pedir um suco de laranja, estará mandando uma mensagem clara.		

Figura 4: Exemplo de resultado do PoNTE compacto, mostrando ao mesmo tempo todas as traduções de uma frase original.

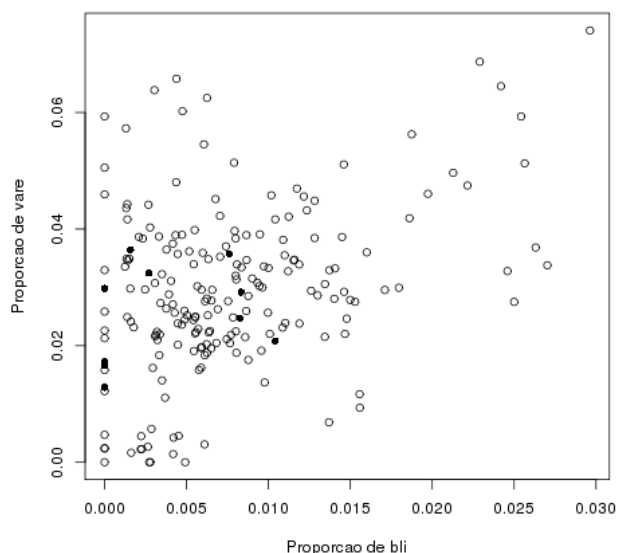


Figura 6: Distribuição da proporção de bli vs. vare em textos originais (bola preta) e traduzidos (bola branca)

Do ponto de vista de onde marcar, parece-me natural marcar associado à unidade de tradução como um todo os casos em que houve dificuldades generalizadas na tradução, enquanto que apenas

faz sentido marcar em cada tradução os casos específicos em que houve um erro ou problema.

Ainda que com o advento de mais traduções de um mesmo texto possa haver casos problemáticos que deixem de o ser (por exemplo, casos em que todas as traduções estavam mal num dado ano podem passar a ter algumas traduções sem problemas ao adicionar as traduções dos anos subsequentes), parece-me relevante distinguir os casos em que a probabilidade de haver problemas (medida pelo número de más traduções) seja maior do que a existência de problemas pontuais, e que esta marcação, por corresponder a mais problemas, tenha prioridade.

Em segundo lugar, qual a “unidade” anotada? A frase inteira? Um sintagma, uma oração? De facto, não existe sempre uma resposta simples sobre o nível (linguístico) em que um determinado erro ou problema deve ser atacado, como aliás o demonstrou cabalmente a tentativa de produzir uma “ontologia” de erros de tradução com o TrAva (Santos, Maia e Sarmento, 2004; Sarmento et al., 2007). Por isso, desde logo consideramos a possibilidade, e vantagem, de um sistema aberto em que um mesmo caso possa ter múltiplas anotações.

Uma coisa é, contudo, clara: a existência de traduções com unidades (frases) relativamente

Linguateca

[Estrutura](#)
[Equipa](#)

[Apresentação](#)
[Acesso a recursos](#)

- [AC/DC](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)
- [COMPARA](#)
- [Corpógrafo](#)
- [CorTrad](#)
- [Esfinge](#)
- [Floresta Sintáctica](#)
- [GeoNetPT](#)
- [PAPEL](#)
- [PoNTE](#)
- [ReEscrire](#)
- [REMBRANDT](#)
- [REPENTINO](#)
- [Repositório](#)
- [WebSpell](#)
- [WPT](#)

[Catálogo de recursos](#)
[Catálogo de ferramentas](#)
[Catálogo de atores](#)
[Catálogo de publicações](#)
[Informação interessante](#)
[Fórum](#)
[Avaliação conjunta](#)
[Perguntas já respondidas](#)

Resultados da pesquisa

Eis os resultados da busca efetuada no PoNTE distribuído.

Procura: [lema="calhar"].
distribuição de , concordância em contexto.
Corpus: PONTEPD

26 ocorrências.

Concordância

Procura: [lema="calhar"].

Num	Original	Traduções	Comentários
1	E quando o Sr. Mário estava mal disposto, o melhor era nem se entrar lá, para não se levar roda do que calhasse.	Og når herr Mario var i dårlig humor, var det best å ikke gå inn der for ikke å gjøre ting verre.	
2	Toda a gente até, se calhar.	Hittil har folk tiet.	
3	E quando o Sr. Mário estava mal disposto, o melhor era nem se entrar lá, para não se levar roda do que calhasse.	Og når herr Mario ikke hadde det så bra, var det best å ikke engang gå inn dit, for ikke å bli kastet ut i rennesteinen.	
4	Toda a gente até, se calhar.	En man ikke motsier.	
5	E quando o Sr. Mário estava mal disposto, o melhor era nem se entrar lá, para não se levar roda do que calhasse.	Og det var best å ikke gå inn der når herre Mário var i dårlig humor, for ikke å få kjeft.	
6	Toda a gente até, se calhar.		
	E quando o Sr. Mário estava mal disposto, o	Og når Herr Mário var i dårlig humor,	

Figura 5: Exemplo de resultado do PoNTE distribuído, mostrando cada tradução separadamente.

extensas leva a que não seja apropriado marcar toda a unidade de tradução quando há um problema que se refere a um vigésimo da mesma. Além disso, essa mesma tradução pode ser digna de elogio em relação a um fenómeno e deficiente em relação a outro.

Para concretizar, veja-se o seguinte exemplo:

- (2) A construção de mais 3.000 unidades, pelo programa «Minha Casa Minha Vida», dará continuidade ao Plano Municipal de Habitação Social no Cidade Aracy e na região do Zavaglia, além da construção de unidades para atender aos servidores públicos municipais.

Bygging av mer enn 3000 enheter, gjennom programmet «Mitt hus mitt hjem», fortsetter den kommunale planen for sosiale boliger i byen Cidade Aracy og i regionen Zavaglia, i tillegg til bygging av enheter for kommunalt ansatte.

A construção de mais de 3000 unidades (...) construção de unidades para empregados da câmara.

Enquanto o uso de um sintagma preposicional simples (*para empregados da Câmara*), em vez de ser fiel à perífrase original, é extremamente comendável, o aprendiz caiu na esparrela de confundir a adição (*mais 3000 unidades*) com a comparação ou com a aproximação (*mais de 3000 unidades*).

Ou seja, na tradução desta frase encontramos um caso exemplar, e um caso de erro, ambos devendo ser analisados e marcados.⁶

Como fazê-lo, porém, levanta a questão das unidades que devem ser o objeto da anotação, e a questão do contexto que deve ser mostrado numa procura (tanto de um lado como do outro), além de como executar essa mesma procura.

Vemos imediatamente que um alinhamento por frase é demasiado grosseiro, e que um alinhamento à palavra também muitas vezes é insuficiente. A própria palavra *alinhamento* é,

⁶Poder-se-ia argumentar que apenas os problemas mereceriam relevo, mas se fosse esse o caso não poderíamos nem usar o PoNTE para dar bons exemplos, nem para fazer estudos quantitativos, porque para esses é preciso anotar tudo (no que se refere ao fenómeno em questão.)

aliás, pouco feliz⁷, visto que queremos pôr em correspondência mas não alinhar.

A última pergunta, mais filosófica, posta pela Belinda Maia aquando da revisão deste artigo, é como escolher os problemas a debater ou focar na aula. Segundo a sua longa experiência, esse é um dos casos em que há menos consenso (no ensino de inglês como língua estrangeira – mas penso que é óbvio que tal problema deve ser extensivo ao ensino de qualquer língua estrangeira). A resposta é que tal é puramente subjetivo, e que depende de cada professor (e sua classe) os casos em que pega na sala de aula. É também relevante distinguir entre ensino de tradução propriamente dito ou ensino da língua. Mas aí penso que o PoNTE poderá ser usado de maneira diferente nos dois tipos de aula, embora não o vá demonstrar em seguida.

O resto desta secção descreve o tipo de comentários que estas traduções suscitam, tentando dar uma ideia dos fenómenos considerados interessantes e usados pedagogicamente.

Divido a discussão entre os casos identificados em cada uma das direções, notando que correspondem a problemas diferentes: enquanto a tradução para norueguês é geralmente feita para a língua materna, a tradução para português corresponde a uma retroversão, e foi sempre corrigida por mim no que se refere a problemas gramaticais, ortográficos ou morfológicos.

5.1 Na direção português-norueguês

Vejam-se alguns problemas que têm mostrado ser geralmente complicados:

Na frase seguinte, em doze traduções, onze estavam erradas no que se refere à estrutura argumental (mostro apenas uma).

- (3) **Defendia** uma vez o meu amigo Nuno, num comentário que aqui deixou, que «as nossas memórias do que foi são muito seletivas».
Jeg forsvarte en gang min venn Nuno i en kommentar som han la ut her, at våre minner om det som var, er høyst selektive.
Eu defendi uma vez o meu amigo Nuno ...

O problema tem a ver com a ordem livre das palavras do português, que – quando comparada com uma língua com uma ordem fixa, como o norueguês – é dificilmente aceite/compreendida pelos alunos, que também aparentemente têm problemas com o facto de a completiva (o objeto) vir tão longe do verbo.

- (4) – quando você terminar o mestrado, a não ser que consiga emprego como pesquisador em empresas privadas (que são *pouquíssimos*), você terá necessariamente que fazer um doutorado?

O problema aqui é que em português é claríssimo, devido ao masculino, que são os empregos que são pouquíssimos, e não as empresas privadas, como a totalidade das traduções permite erradamente inferir – porque não há género nem número de adjetivos em norueguês.

- (5) – então, *com 3 anos de formado*, você terá que concorrer a bolsas de R\$ 2.000 mensais para fazer doutorado?

Nesta frase, *com 3 anos de formado* significa que já passaram 3 anos desde que se formou... mas esta construção não era conhecida dos alunos, e por isso ninguém a conseguiu interpretar (e consequentemente traduzir) corretamente.

Outro problema bem distinto:

- (6) Encararam e puseram o problema de fundir, adoptar soluções de compromisso ou separar radicalmente culturas por vezes altamente complexas (a Indiana, a Chinesa, as Africanas, a Brasileira) e religiões (Budismo, Bramanismo).

Possivelmente assoberbados com tanta informação numa frase, em (6) nenhum aluno conseguiu compreender que havia uma escolha entre três formas diferentes de proceder (nomeadamente (i) fundir, ou (ii) adotar soluções de compromisso, ou (iii) separar radicalmente), tendo portanto traduzido por algo muito pouco claro.

Outros casos são lexicais e têm a ver com diferentes culturas. No caso seguinte, é preciso compreender que as baratas não têm o mesmo papel de representantes do submundo em Moçambique e na Noruega:

- (7) Mas o que dói saber é que, por extensão e acumulação, o não-funcionamento de tudo acabaria por servir aqueles que, como os ratos e as baratas, se movem melhor no caos e na podridão.

ou compreender o que significa “despachar”:

- (8) Despachar um documento preso nas burocracias;
gjøre kort prosess med et dokument som er fanget i byråkratiet;
Å sende et fast dokument innenfor byrokra-

⁷Como já argumentado em Santos e Simões (2008).

tiet.

Avsende et dokument som var fastklemt i
byråkratiet
Ekspedere et dokument som er fanget i
byråkratiet;
avlevere et document som står fast i
byråkratiet

É muito interessante constatar que quatro tradutores interpretaram *despachar* como enviar pelo correio, e um como liquidar ou contornar.

De qualquer forma, o caso mais complicado é a que costume dar mais atenção tem a ver com a estruturação do discurso. Muitas vezes essa não é compreendida, e daí faltar coesão, por vezes mesmo qualquer lógica, à tradução resultante. De seguida, apresento alguns exemplos diversos de casos que tendem a ser ignorados:

- (9) **E se** os Portugueses foram ajudados por inúmera gente de muitos países e tradições, não resta dúvida de que o esforço de aquisição foi seu, como sua foi a consciência primeira do novo mundo e o desafio àquele que existia.
- (10) As mercearias são **o menos**.
- (11) A contribuição de Portugal para o Renascimento, **todavia**, não se deu tanto no capítulo das Artes ou das Humanidades como no da Ciência.
- (12) **Já** a solução do problema da dívida deve ser combinada com o crescimento econômico.
- (13) Feitas as ressalvas, **vamos então** à minha campanha de anti-propaganda sobre a ciência no Brasil!

Outros comentários são mais subjetivos, e dependem naturalmente do estilo e do gosto, mas também do “espírito da língua”. No caso seguinte, é natural comentar como “Desnecessário” traduzir uma oração participial (que não existe em norueguês) por uma relativa:

- (14) A mesma comida, a mesma decoração, o mesmo programa na televisão **pendurada na parede**, ouvindo o atendente pronunciar a mesma palavra-chave: «hambúguer».
Den samme maten, den samme dekorasjon og det samme TV-programmet på TV-en **som er hengt på veggen**, og kelneren som sier det samme nøkkelordet: «hamburger».
‘que está pendurada na parede’

Os casos mais interessantes – mas certamente

mais difíceis de anotar – são aqueles em que mais de uma diferença contrastiva se conjuga para complicar a expressão do sentido original, veja-se por exemplo:

- (15) Embora não haja um número expressivo de dados sobre os quais apoiar as diferenças, **qualquer observador mais atento** é capaz de verificar que elas existem e talvez até permitam identificar a que região da cidade um carioca pertence.

Neste caso, *qualquer observador mais atento* não foi bem traduzido/compreendido por nenhum dos 12 tradutores, possivelmente pela dificuldade conjugada (i) da palavra *observador* (o que observa e não necessariamente um papel/profissão), (ii) do uso da comparação sem termo de comparação, e (iii) do emprego do quantificador *qualquer*, que por ser intensional é sempre difícil de exprimir numa língua germânica.

- (16) Queremos – e **podemos** – ajudar, enquanto há tempo, os países onde a crise já é aguda.

Neste caso, é preciso atentar à variada polissemia de *poder*, e escolher a mais adequada e não a mais frequente. No exemplo acima, *podemos* está empregue no sentido de “temos meios/capacidade para” e não no sentido de mera possibilidade.

E ainda mais difícil é quando é preciso traduzir formas mais criativas⁸:

- (17) Eis que, de modo impensado, o subdesenvolvimento cria o seu próprio vocabulário, **as suas formas de se dizer**.

As suas formas de se dizer é certamente uma expressão que merece, mesmo para falantes nativos, alguma reflexão (basta ser de um escritor tão extraordinário como Mia Couto). Mas textos que valham a pena ser traduzidos são precisamente textos que provoquem reflexão.

Estes exemplos demonstram, espero, que a variedade de problemas ou de questões merecedoras de comentário, entrando apenas em conta com o texto fonte, é vastíssima, mas inescapável.

5.2 Na direção norueguês-português

Tendo em conta o objetivo destas traduções, ou retroversões, na direção norueguês-português,

⁸Note-se que não me estou a referir ao caso, também frequente, mas mais fácil de detetar, de formas elas próprias inventadas e portanto novas, como *brinciar* no texto em questão.

o tipo de comentários e de problemas é naturalmente outro, e está centrado no mantra do “não é assim que se exprime isso em português”, no espírito por exemplo de (Bennett, 2010), ou então, no alertar para ter cuidado com outras interpretações que a tradução erroneamente adicionou.

Por exemplo, a questão da posição e forma do movimento, e dos modais, são críticas: na maior parte das vezes a primeira deve omitir-se, e a segunda traduzir-se pelo tempo e modo correspondentes.

Veja-se um primeiro exemplo em que a posição seria completamente irrelevante em português:⁹

- (18) Her **sitter** jeg og faar ikke sove fordi jeg er saa fortvilt.
 Estou sentado aqui e não consigo dormir porque estou desesperado.
 Agora estou tão desesperada que não consigo dormir,
 Estou aqui, sem poder dormir, porque me sinto muito desesperada.
 Aqui estou eu, tão preocupada que não consigo dormir.
 Estou sentada e não consigo dormir por causa do desespero.
 Estou sentado aqui sem conseguir dormir porque eu estou muito desesperado.

Um caso trivial em que o verbo modal nunca deveria ser traduzido literalmente é o seguinte.

- (19) Hvorfor sauer ikke **kan** svømme
 Porque as ovelhas não se põem a nadar.
 Porque ovelhas não sabem nadar
 Porque os carneiros não nadam
 Porque os carneiros não sabem nadar?
 Por que as ovelhas não nadam
 Porque é que as ovelhas não sabem nadar
 Porque é que os carneiros não nadam?
 Porque ovelhas não sabem nadar
 Porque as ovelhas não sabem nadar
 Por que é que as ovelhas não sabem nadar.
 Porque os carneiros não sabem nadar
 Porque carneiros não sabem nadar
 Porque as ovelhas não são capazes de nadar.

Repare-se que a variação entre *carneiros* e *ovelhas* é completamente irrelevante, neste caso, mas não a distinção entre *porquê* e *porque* (*hvorfor* e *fordi* em norueguês), em que quase nenhum tradutor reparou, e que fundamentaria, em minha opinião, algo tão diferente como *O porquê dos carneiros*

não nadarem...

Noutros casos é simplesmente o tom, ou o tipo de língua, que não é conseguido pelas várias traduções. O exemplo seguinte é também interessante pela transformação de *der* (onde) para *quando* ou *enquanto*, necessária em português.

- (20) Siss hadde mange tankar **der** ho gjekk, innballa for frosten.
 Muitas coisas ocupavam a cabeça de Siss, que estava andando no frio ...
 Siss tinha muitos pensamentos andando por ali, embalada contra o frio.
 Siss tinha muitos pensamentos enquanto caminhava, coberta contra o frio
 Siss tinha vários pensamentos, envolvida devido à geada.
 Siss está a pensar, enrodilhada por causa do frio:
 Siss estava pensando muito quando caminhava pela floresta, embrulhada em roupas contra o frio.

As questões discursivas são igualmente relevantes nesta direção da tradução, claro, e embora no que se segue eu apresente todas as traduções para dar uma ideia do conteúdo e da necessidade ou falta da sua tradução, essa ideia resulta bastante pálida porque a maior parte destes marcadores não foi bem traduzida:

- (21) Jo mer overraskende resultatet er, jo bedre er forskningen, **på et vis**.
 Quanto mais surpreendente o resultado, melhor a pesquisa, tipo.
 Quanto mais estranhos os resultados, melhor a investigação... de uma certa forma.
 Quanto mais surpreendente o resultado, melhor a investigação, por um lado.
 Quanto mais surpreendente o resultado, melhor a pesquisa, de alguma forma.
 De certa forma, mais surpreendente o resultado melhor a pesquisa.
 O fato é que as pesquisas mais surpreendentes são as melhores, mas às vezes uma pesquisa torna-se mais surpreendente de que originalmente era para ser.
 Duma maneira, quanto mais surpreendente for o resultado, melhor é a pesquisa.
 De um jeito, quanto mais surpreendente for o resultado, tanto melhor para a investigação.
 Tanto mais surpreendente o resultado, tanto melhor a investigação.
 Quanto mais surpreendente o resultado, melhor a pesquisa, de alguma maneira.

⁹Na direção do norueguês para o português apresento muitas traduções, para ilustrar o nível do português dos alunos e a divergência de traduções no PoNTE.

E quanto mais surpreendente o resultado, tanto melhor a pesquisa, de algum modo.

- (22) Noen ganger høres **nemlig** ting tullete ut fordi de er tullete.

Às vezes as coisas parecem ridículas exatamente porque são ridículas.

Às vezes as notícias parecem disparatadas, porque são disparatadas.

Algumas vezes as coisas parecem tolas porque o são.

Às vezes as coisas parecem tolas porque o são.

Às vezes, coisas que parecem absurdas, realmente são absurdas.

Algumas vezes as coisas soam disparatadas porque são disparatadas.

Algumas vezes coisas parecem desatinadas porque realmente são.

De vez em quando as coisas parecem asneiras porque é a característica correta.

Às vezes mesmo coisas parecem brincadeiras porque são brincadeiras.

É que, às vezes, uma coisa parece tola porque é tola.

- (23) Men jeg er **ganske sikker** på at de regnet med at jeg drakk som dem. Når jeg **faktisk** lurte dem.

Mas tenho quase a certeza de que pensavam que eu estava a beber também, quando **realmente** os estava a enganar.

Mas tenho certeza que eles estavam contando com que eu estivesse bebendo álcool assim como eles, quando **na verdade** eu os enganei.

- (24) Nogle af potterne er desværre knækket, men det kan man **næppe** undgå efter en forsendelse fra Sydamerika.

Infelizmente, alguns dos potes estão quebrados, mas isto não é possível evitar quando se trata de um envio da América do Sul.

Infelizmente alguns dos jarros estão partidos, mas isto é quase impossível de evitar num envio da América do Sul.

Alguns dos vasos estão infelizmente quebrados, mas isto é quase inevitável depois de uma viagem proveniente da América do Sul.

5.3 Pilotos de anotação

Para avançar na determinação da forma mais prática de adicionar a anotação, desenvolvi um programa auxiliar que permitia anotar informação sobre as traduções de dados verbos (qual verbo/lema era um parâmetro do programa), e criei um novo atributo posicional

(coluna) no PoNTE distribuído, de forma a poder ter uma ideia da gama (e distribuição) dos diferentes comentários.

Anotei depois todas as ocorrências dos verbos *haver*, *ser*, *estar*, *ficar* e *ter*, que passaram a ser procuráveis pela interface do PoNTE distribuído, pedindo a distribuição de traduções, como ilustrado na figura 7.¹⁰

Ao fazer isso, logo várias questões surgiram, como os leitores pela simples leitura do exemplo podem apreciar:

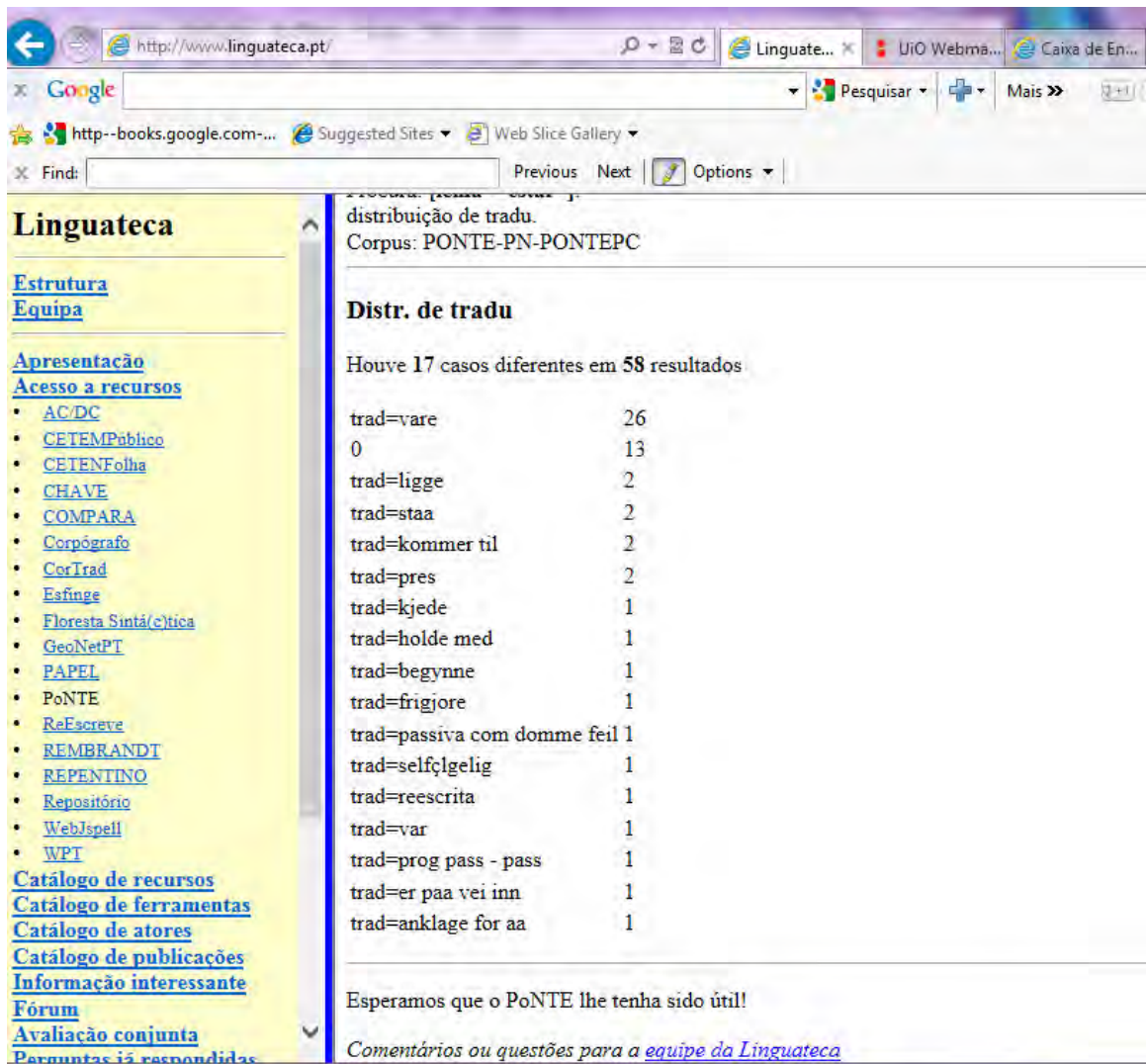
1. Que tipo de valores associar?
2. Como padronizar os comentários?
3. A que nível de detalhe descer?
4. Como distinguir as duas situações diversas: (i) não traduzido porque o aluno simplesmente não fez de (ii) não traduzido porque o tradutor escolheu outra forma?
5. Será que associar as traduções às palavras é uma boa ideia, ou deveria ser a uma unidade maior?
6. E: Faz sentido anotar todas as palavras, ou apenas aquelas que nos suscitarem comentários?

E, talvez o problema maior de todos, imediatamente ilustrado pela nota de rodapé relativa aos “0”, como garantir uma anotação atualizada, quando a adição de novas traduções transforma os dados quantitativos anteriormente adicionados em valores incorretos?

Todas estas perguntas correspondem a decisões que terão de ser tomadas, possivelmente com experimentação de vários caminhos e perguntas a vários utilizadores.

Outra anotação foi também efetuada no âmbito do estudo da tradução específica de sentimentos e sensações, agora classificando simplesmente os verbos (em português) em várias categorias sintático-semânticas (veja-se Santos (2014e) para mais pormenores). Aí ficou outra vez patente a dificuldade de refazer os corpos com mais informação, dado existirem diferentes fontes de anotação que é preciso harmonizar para cada nova versão. Ou seja, se da versão 3.0 para a 4.0 se passa a adicionar anotação automática de partes do corpo humano, como reintroduzir a anotação humana das sensações aqui mencionada, se ficou guardada em versões anteriores sem partes do corpo?

¹⁰Os casos de 0 correspondem a casos de ocorrências desses verbos que não foram anotadas, porque pertencem a traduções que foram incorporadas depois do piloto ter tido lugar.



The screenshot shows a web browser window with the URL <http://www.linguateca.pt/>. The page title is "Linguateca". On the left, there is a sidebar with a menu containing links like "Estrutura", "Equipa", "Apresentação", "Acesso a recursos", "Catálogo de recursos", "Catálogo de ferramentas", "Catálogo de atores", "Catálogo de publicações", "Informação interessante", "Fórum", "Avaliação conjunta", and "Perguntas já respondidas". The main content area shows search results for "Distr. de tradu". It states "Houve 17 casos diferentes em 58 resultados". Below this, there is a table listing various translation cases and their frequencies.

Distr. de tradu	
Houve 17 casos diferentes em 58 resultados	
trad=vare	26
0	13
trad=ligge	2
trad=staa	2
trad=kommer til	2
trad=pres	2
trad=kjede	1
trad=holde med	1
trad=begynne	1
trad=frigjore	1
trad=passiva com domme feil	1
trad=selfølgelig	1
trad=reescrita	1
trad=var	1
trad=prog pass - pass	1
trad=er paa vei inn	1
trad=anklage for aa	1

Below the table, there is a message: "Esperamos que o PoNTE lhe tenha sido útil!" and a link: "Comentários ou questões para a [equipe da Linguateca](#)".

Figura 7: Um exemplo de comentários às traduções.

Estas questões levam a que seja mais natural tentar manter essa informação separadamente, para poder ser reposta numa nova versão, mas também isso não é necessariamente à prova de problemas entre versões, visto que pode haver diferenças de atomização ou outras, que inviabilizem uma mesma identificação... como tivemos experiência disso no caso de dois corpos que incorporámos no AC/DC mas que já vinham com anotação criada por métodos diferentes: a CDHAREM (Rocha e Santos, 2007) e o ReLi (Freitas et al., 2012).

Uma possibilidade é, naturalmente, transformarmos o processo de anotação em algo automático que é repetido para cada nova versão do corpo, como foi feito no caso do piloto do tipo de emoções, usando o *corte-e-costura* (Mota e Santos, 2009; Santos e Mota, 2010). Isso implica criar regras específicas, e usá-las de novo sempre que se recria o corpo. O problema é que se o corpo aumentou (incluindo novos textos, por

exemplo), as regras muito provavelmente têm de ser alargadas para cobrir o novo material – ou então podem resultar mais prejudiciais do que úteis, se quisermos fazer fé nos dados quantitativos.

Parece que não podemos senão concluir que versões diferentes de um corpo em constante desenvolvimento terão de incluir questões diferentes, e que, para ter um produto acabado, não podemos melhorar noutros campos, o que é, de facto, na minha opinião um dos dilemas maiores dos compiladores de corpos:

- Dizemos que o corpo está pronto, estável, e não se mexe mais?
- Ou, pelo contrário, podemos/devemos ir melhorando e adicionando mais informação e mais textos?

Um sistema de versões para corpos é demasiado pesado, e a maior parte dos utilizadores nem sequer cita a versão (nem a data de acesso),

quanto mais compara com as anteriores...É a esse respeito instrutivo comparar o que escrevi em 2000 sobre o CETEMPúblico (Santos, 2000) e o que concluí mais tarde, nomeadamente que essas expectativas eram completamente desajustadas da realidade (Santos, 2014a).

Não tenho respostas definitivas para estas perguntas, mas devo mencionar que é pelo menos lícito fazê-las...e que estamos a voltar à carga neste assunto (a reutilização de anotações) com a iniciativa da Gramateca¹¹, lançada em janeiro do presente ano de 2014.

6 Comentários finais

Espero ter demonstrado que a avaliação de traduções na aprendizagem de uma língua é um fenómeno complexo que merece o desenvolvimento de ferramentas apropriadas ao seu estudo e ao seu reuso, e que um corpo como o PoNTE, dotado das ferramentas idealizadas acima, pode ser útil a dois níveis:

- a nível pedagógico, para ensinar este tipo de aprendentes e para formar futuros professores na área (e no par de línguas em questão);
- a nível gramatical, eventualmente para estudar outros tipos de fenómenos gramaticais que sejam suficientemente frequentes para permitir a obtenção de vários exemplos, bem e mal traduzidos.

Contudo, a implementação de um sistema que permita de facto procurar e codificar o tipo de informação desejada não é obviamente uma tarefa simples, e ainda nos encontramos nos primórdios de tal realização.

Que eu conheça, apenas Oliveira (2012) publicou um estudo que usa traduções de alunos num corpo paralelo que incluía o português. Kutuzov et al. (2012), por outro lado, é um projeto semelhante ao PoNTE em termos de abrangência e objetivos (para o par russo e inglês), mas com diferentes soluções técnicas e aparentemente ainda sem uso no próprio ensino.

Embora com espírito semelhante, nem Bernardino (2002) nem Abekawa e Kageura (2008) ou Bojar et al. (2008), que relatam trabalhos ou ideias interessantes associadas à anotação de corpos paralelos, têm um recurso do mesmo tipo que o PoNTE.

Embora existam alguns artigos relacionados com o ensino da tradução e/ou o estudo da atividade dos aprendizes de tradutores envolvendo o português como uma das línguas, não consegui identificar mais nenhum sobre um corpo de trabalhos produzidos por este tipo de aprendentes, o que me leva a insistir que este tipo de corpo é escasso e inovador.

Pela minha própria pesquisa de referências a trabalhos parecidos, é certamente Popescu-Belis, King e Bentanar (2002) aquele que é mais consonante com o descrito aqui, visto que os autores descrevem um corpo de traduções corrigidas, no âmbito do ensino (e avaliação) da tradução. Mas as semelhanças acabam aí: O formato escolhido é XML, e usam a nota atribuída a cada tradução pelos examinadores como (uma das) formas de avaliar cada tradução globalmente. Além disso, parece-me ser mais a correção da língua de chegada que está em questão na correção das traduções e não tanto a própria tradução. Finalmente, o par é francês para inglês; o conteúdo, em 2002, correspondia a 50 traduções de dois textos, e um dos objetivos da criação do corpo era para servir de treino na avaliação de tradução automática.

Não quero contudo dar a entender com as afirmações anteriores que não exista qualquer trabalho que estude aprendentes ou aprendizes de língua portuguesa, ou a aprender português:

- existem vários chamados “corpora de aprendiz de língua estrangeira”, por exemplo Shepherd (2009), Dutra e Silero (2012) e Tagnin e Fromm (2008) tratam de corpos de alunos brasileiros a aprender inglês (no caso do COMAprend, também francês, alemão, italiano e espanhol), e Gamallo et al. (2013) refere-se a alunos portugueses a aprender galego;
- enquanto que Evers e Wilkens (2012) e os corpos PEAPL2¹² da Universidade de Coimbra e *Recolha de Dados de Aprendizagem de Português Língua Estrangeira*¹³, do Centro de Linguística da Universidade de Lisboa, contêm textos de aprendizes de português como língua estrangeira.

A maior diferença é que nenhum destes corpos tem origem em traduções, mas sim em atividades de redação ou interação na língua a que se referem. Além disso, e pelo que me foi dado apreciar, a maioria também não tem

¹¹Ver <http://www.linguatca.pt/Gramateca>, veja-se uma primeira apresentação do projeto em Santos (2014b; Santos (2014c).

¹²<http://www.uc.pt/fluc/rcpl2/>

¹³<http://www.clul.ul.pt/pt/recursos/314-corpora-of-ple>

associado qualquer sistema de procura específico, ao contrário do PoNTE.

Mas é inegável que a combinação de todos estas peças de um mesmo quebra-cabeças é interessante para estudar as dificuldades da aprendizagem do português e as características do português que diferem de outras línguas, e que um trabalho futuro pertinente seria desenhar estudos que usassem os três tipos de materiais.

Este artigo é, contudo, apenas dedicado ao PoNTE. Embora seja trabalho em progresso, já nos parece merecedor de publicação, tanto para congregar futuros utilizadores como para pôr à consideração e discussão do público em geral possíveis caminhos a seguir num futuro próximo, nomeadamente:

- a criação de um novo “corpo” com apenas anotações, alinhado com os corpos de texto a que se refere, em vez de incluir as anotações como atributos dos corpos;
- a expansão do Ensinador (Simões e Santos, 2011) para corpos paralelos, em que ao aluno são apresentados alternativas relevantes para a própria tradução (desde que essa seja considerada correta), semelhante ao que é feito no RuN (Grønn e Marijanovic, 2010);
- o desenvolvimento de um programa que permita identificar automaticamente os casos de problemas, por exemplo comparando as várias traduções, o que é algo também relevante para estudos mais profundos do processo de tradução como os que pretendemos fazer no âmbito do CorTrad.

Gostaria contudo de terminar este texto com duas notas negativas, propostas pela Belinda Maia, e que me parecem importantes para diminuir um exagerado otimismo que possa ficar na imaginação dos leitores.

Em primeiro lugar, não é garantido que a crítica de traduções a alunos de língua não seja contraproducente. Repare-se que os alunos não são alunos de tradução, mas sim de língua. Uma coisa é achar que lhes torna as aulas mais interessantes – e a mim também, a outra é realmente conseguir demonstrar que aprendem melhor a língua portuguesa (e/ou que eu a ensino melhor).

Em segundo lugar, não é garantido que o trabalho – relativamente grande – de criar este corpo seja rentável em termos práticos (comparando com outras tarefas que eu poderia fazer para a melhoria do ensino), sobretudo devido à falta de utilizadores neste par de línguas. Só o futuro, realmente, o dirá. Futuro esse

que, nos tempos mais próximos, pode ir sendo auscultado pelo ritmo de mudança e melhoria do corpo.

Agradecimentos

O trabalho aqui descrito enquadra-se no âmbito da Linguatca, co-financiada desde o seu início pelo Governo Português, pela União Europeia (FEDER e FSE), sob o contrato POSC/339/1.3/C/NAC, pela UMIC e pela FCCN; de 2009 até 31 de dezembro de 2011 pela Fundação da Ciência e da Tecnologia (FCT) e pela Fundação para a Computação Científica Nacional (FCCN), e a partir dessa data apenas apoiada pelas Universidades em que os membros trabalham.

Especificamente o projeto PoNTE foi apoiado pela Universidade de Oslo através da atribuição, no Outono de 2011, de uma mini-bolsa de investigação a Marcin Wlodek, a quem estou grata pelos comentários relativos às traduções desse semestre.

Agradeço também a Joacyr Oliveira as discussões sobre o uso de corpos de aprendentes de tradução para ensinar a língua e a tradução, e a Belinda Maia os comentários pertinentes na sua recensão.

Referências

- Abekawa, Takeshi e Kyo Kageura. 2008. Constructing a corpus that indicates patterns of modification between draft and final translations by human translators. Em *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. European Language Resources Association (ELRA), 28-30 Maio, 2008.
- Bakkejord, Kaja Rindal. 2008. Técnicas de substituição e supressão dos clíticos no português do Brasil. Tese de Mestrado, Universidade de Oslo.
- Bennett, Karen. 2010. Academic discourse in portugul: A whole different ballgame? *Journal of English for Academic Purposes*, 9(1):21–32.
- Bernardini, Silvia. 2002. Educating translators for the challenges of the new millenium: The potential of parallel bidirectional corpora. Em Belinda Maia, Johann Haller, e Margherita Ulrych, editores, *Training the language services provider for the New Millenium, Proceedings of the III Encontros de Tradução*. Astra-FLUP, FLUP, Porto, pp. 173–186.

- Bick, Eckhard. 2000. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Tese de doutoramento, Aarhus University, Aarhus, Denmark, November, 2000.
- Bojar, Ondrej, Miroslav Janicek, Zdenek Zabo-
krtsky, Pavel Ceska, e Peter Bena. 2008. CzEng 0.7: Parallel Corpus with Community-Supplied Translations. Em *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. European Language Resources Association (ELRA), 28-30 de maio, 2008.
- Borin, Lars, Lauri Carlson, e Diana Santos. 2001. Corpus based language technology for computer-assisted learning of Nordic languages: Squirrel. Em Henrik Holmboe, editor, *Nordisk sprogteknolog (Nordic language technology)*, *Aarvog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. Museum Tusculanum Forlag, Københavns Universitet, Copenhagen, pp. 257–270, Setembro, 2001.
- Costa, Luís, Diana Santos, e Paulo Alexandre Rocha. 2009. Estudando o português tal como é usado: o serviço AC/DC. Em *The 7th Brazilian Symposium in Information and Human Language Technology (STIL 2009)*, 8-11 de setembro, 2009.
- Dutra, Deise Prina e R. P. Silero. 2012. O uso de for: uma análise de itens linguísticos em corpus de aprendizes brasileiros. Em Tania Shepherd, Tony Berber Sardinha, e Marcia Veirano Pinto, editores, *Caminhos da linguística de corpus*, pp. 325–341. Mercado de Letras.
- Evers, Aline e Rodrigo Wilkens. 2012. Classificação de proficiência em língua adicional no português um estudo para a determinação de índices diferenciadores. Em *IX Encontro Nacional de Inteligência Artificial, ENIA 2012*, 20-25 de outubro, 2012.
- Frankenberg-Garcia, Ana e Diana Santos. 2002. COMPARA, um corpus paralelo de português e de inglês na Web. *Cadernos de Tradução*, IX(1):61–79.
- Freitas, Cláudia, Eduardo Motta, Ruy Luiz Milidiú, e Juliana César. 2012. Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. Em *XI Encontro de Linguística de Corpus - ELC 2012*, 13-15 de setembro, 2012.
- Gamallo, Pablo, Marcos García, I. González, Muñoz. M., e I. Del Río. 2013. An evaluation of Avalingua based on learner corpora. Em *ICAME34 Workshop Learner Corpora and their Application in Language Testing and Assessment, May 22, Santiago de Compostela, Spain, 2013*, pp. 52–53.
- Granger, Sylviane, Gaëtanelle Gilquin, e Fanny Meunier. 2013. *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)*. Presses universitaires de Louvain.
- Grønn, Atle e Irena Marijanovic, 2010. *Russian in contrast: form, meaning and parallel corpora*, pp. 1–24. Oslo Studies in Language 2(1).
- Kutuzov, A. B., M. A. Kunilovskaya, A. Y. Oschepkov, e A. Y. Chepurkova. 2012. Russian-learner parallel corpus as a tool for translation studies. Em *Proceedings of Dialog 2012*.
- Maia, Belinda e Diana Santos. 2012. Who is afraid of ... what? - In English and in Portuguese. Em Signe Oksefjell Ebeling, Jarle Ebeling, e Hilde Hasselgård, editores, *Aspects of corpus linguistics: compilation, annotation, analysis*, number 12 in *Studies in Variation, Contact and Change in English*, Dezembro, 2012.
- Mota, Cristina e Diana Santos. 2009. Corte e costura no AC/DC: auxiliando a melhoria da anotação nos corpos, Setembro, 2009. <http://www.linguateca.pt/acesso/corte-e-costura.pdf>.
- Nilsson, Kåre. 1997. A lusofonia vista por um lusitanista escandinavo, 11 de novembro, 1997. 1.o encontro de Professores de Português - Língua Estrangeira, organizado pelo Centro de Línguas, FFLCH, USP.
- Oliveira, Joacyr. 2012. A linguística de corpus na formação de tradutores: compilação e análise de um corpus de aprendizes de tradução. Em *Trabalho em andamento - Anais do XI Encontro de Linguística de Corpus (ELC 2012)*.
- Popescu-Belis, Andrei, Margaret King, e Houcine Bentanar. 2002. Towards a corpus of corrected human translations. Em Margaret King, editor, *Machine Translation Evaluation - Human Evaluators Meet Automated Metrics, Workshop Proceedings, LREC2002*, pp. 17–21.
- Rocha, Paulo e Diana Santos. 2007. Disponibilizando a <OBRA>

- Colecção Dourada </OBRA> do <ACONTECIMENTO > HAREM </ACONTECIMENTO> através do projecto <LOCAL|ORGANIZACAO|ABSTRACCAO> AC/DC </LOCAL|ORGANIZACAO|ABSTRACCAO>. Em Diana Santos e Nuno Cardoso, editores, *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*, pp. 307–326. Linguateca, 12 de novembro, 2007.
- Rosa, Alexandra Assis. 2006. Does translation have a say in the history of our contemporary linguacultures? Some figures on translation in Portugal. *Polifonia*, 9:77–94.
- Santos, Diana. 2000. O projecto Processamento Computacional do Português: Balanço e perspectivas. Em Maria das Graças Volpe Nunes, editor, *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, pp. 105–113, São Paulo, 19-22 de novembro, 2000. ICMC/USP.
- Santos, Diana. 2002. DISPARA, a system for distributing parallel corpora on the Web. Em Nuno Mamede e Elisabete Ranchhod, editores, *Advances in Natural Language Processing (PorTAL 2002)*, Lecture Notes in Artificial Intelligence, pp. 209–218, Berlin/Heidelberg, 23-26 de junho, 2002. Springer-Verlag.
- Santos, Diana. 2003. Timber! Issues in treebank building and use. Em Jorge Baptista, Isabel Trancoso, Maria das Graças Volpe Nunes, e Nuno J. Mamede, editores, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003. Faro, Portugal, June 2003*, pp. 151–158, Berlin/Heidelberg. Springer Verlag.
- Santos, Diana. 2012. Os possessivos estão-me a complicar o ensino :-), 26 de outubro, 2012. <http://www.linguateca.pt/Diana/download/posterAPL2012.pdf>.
- Santos, Diana. 2013. Ser, estar, ficar, haver and ter against ha, bli and vare: who said it was easy to describe feelings and sensations?, 15 de maio, 2013. <http://www.linguateca.pt/Diana/download/KKPonte.pdf>.
- Santos, Diana. 2014a. Corpora at Linguateca: vision and roads taken. Em Tony Berber Sardinha e Telma São Bento Ferreira, editores, *Working with Portuguese corpora*, pp. 219–236. Bloomsbury.
- Santos, Diana. 2014b. First steps of Gramateca: a corpus-based grammar initiative for Portuguese, driven by Linguateca, 20 de fevereiro, 2014. <http://www.linguateca.pt/Diana/download/Gramateca0slo.pdf>.
- Santos, Diana. 2014c. Gramateca: corpus-based grammar of Portuguese. Em Jorge Baptista, Nuno Mamede, Sara Candeias, Ivandré Paraboni, Thiago A.S. Pardo, e Maria das Graças Volpe Nunes, editores, *PROPOR 2014*, pp. 214–219. Springer Verlag, outubro, 2014.
- Santos, Diana. 2014d. Os possessivos estão-me a complicar o ensino :-). Um estudo do dativo possessivo baseado em corpos. Em apreciação, versão preliminar, <http://www.linguateca.pt/Diana/download/PossAprec.pdf>.
- Santos, Diana. 2014e. Ser, estar, ficar, haver e ter vs. ha, bli e vare: quem disse que era fácil traduzir sentimentos e sensações? Em Signe Oksefjell Ebeling, Atle Grønn, Kjetil Rå Hauge, e Diana Santos, editores, *Corpus-based Studies in Contrastive Linguistics*, pp. 271–288. Oslo Studies in Language 6(1).
- Santos, Diana, Belinda Maia, e Luís Sarmento. 2004. Gathering empirical data to evaluate MT from English to Portuguese. Em Lambros Kranias, Nicoletta Calzolari, Gregor Thurmair, Yorick Wilks, Eduard Hovy, Gudrun Magnusdottir, Anna Samiotou, e Khalid Choukri, editores, *Proceedings of LREC 2004 Workshop on the Amazing Utility of Parallel and Comparable Corpora*, pp. 14–17, 25 de maio, 2004.
- Santos, Diana e Cristina Mota. 2010. Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora. Em Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, e Daniel Tapias, editores, *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)*, pp. 1437–1444. European Language Resources Association, 17-23 de maio, 2010.
- Santos, Diana e Alberto Simões. 2008. Portuguese-English word alignment: some experiments. Em *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. European Language Resources Association (ELRA), 28-30 de maio, 2008.
- Sarmiento, Luís, Anabela Barreiro, Belinda Maia, e Diana Santos. 2007. Avaliação

- de Tradução Automática: alguns conceitos e reflexões. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*, pp. 181–190, Lisboa, Portugal, 20 de março, 2007. IST Press.
- Shepherd, Tania. 2009. Corpora de aprendiz de língua estrangeira: um estudo contrastivo de n-gramas. *Veredas*, 11(2):100–116.
- Simões, Alberto e Diana Santos. 2011. Ensinador: corpus-based Portuguese grammar exercises. *Procesamiento del Lenguaje Natural*, 47:301–309, Setembro, 2011.
- Tagnin, Stella Esther Ortweiler e Guilherme Fromm. 2008. COMAprend – a experiência da construção de um corpus de aprendizes para estudos. *Domínios de Linguagem*, 2(2).
- Tagnin, Stella O. E., Elisa Duarte Teixeira, e Diana Santos. 2009. CorTrad: a multiversion translation corpus for the Portuguese-English pair. *Arena Romanistica*, 4:314–323.
- Teixeira, Elisa D., Diana Santos, e Stella E. O. Tagnin. 2011. CorTrad: um novo corpus paralelo multiversão para o par de línguas português-inglês. Em Tania Shepherd, Tony Berber Sardinha, e Marcia Veirano Pinto, editores, *Caminhos na Linguística de Corpus*. Mercado de Letras, pp. 151–176.